

Im Original veröffentlicht als:

Bertelmann, R. Vom Dokumentenserver zum Institutional Repository In: Erfolgreiches Management von Bibliotheken und Informationseinrichtungen Ed.: Hobohm, C.; Umlauf, K. Hamburg : Dashöfer 2006. 9.4.3. [12] p.

## **Vom Dokumentenserver zum Institutional Repository**

**Roland Bertelmann**

Schon 2001 wurde in Empfehlungen des Wissenschaftsrats gefordert, dass sich Wissenschaftler beim Abschluss von Verlagsverträgen ein entsprechendes weiteres Nutzungsrecht sichern und die von ihnen verfassten Dokumente zusätzlich auf einem Dokumentenserver veröffentlichen sollen.

### **Offener Zugang**

Im Herbst 2003 wurde dann die Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen verabschiedet, die heute weltweit die Unterstützung durch zahlreiche Wissenschaftsinstitutionen und Wissenschaftsorganisationen gefunden hat. Alle Wissenschaftsorganisationen Deutschlands gehören zu den Erstunterzeichnern. In der Berliner Erklärung wird gefordert: „Eine vollständige Fassung der Veröffentlichung wird ... auf mindestens einem online zugänglichen Archivserver mit geeigneten technischen Standards (wie die von Open Archive) hinterlegt und damit veröffentlicht. Der Archivserver muss betrieben werden von einer wissenschaftlichen Institution oder Gesellschaft, einer öffentlichen Institution oder einer anderen etablierten Organisation, die das ‚Prinzip des offenen Zugangs‘, uneingeschränkte Verbreitung, Interoperabilität und Langzeitarchivierung zu verwirklichen sucht.“

Im Januar 2005 empfahlen Senat und Hauptausschuss der DFG geförderte Wissenschaftler aufzufordern ihre Forschungsergebnisse auch offen zugänglich zu machen und die Verwendungsrichtlinien entsprechend zu ändern.

### **Definition**

Der Begriff Institutional Repository (IR) setzt sich im Umfeld akademischer Informationsversorgung zunehmend für diese Art von Diensten durch. Institutional Repositories werden zum unverzichtbaren Bestandteil des Serviceangebots einer Bibliothek und stellen beim Management der EDV-Anwendungen einer Bibliothek neue Anforderungen.

Eine pauschale Definition für ein IR könnte so lauten: ein elektronisches System, in dem digitale Objekte einer Institution oder einer wissenschaftlichen Community gesammelt und archiviert werden und das den Zugang zu diesen Objekten ermöglicht. (nach Foster, Gibbons 2005).

Also ein stabiler Speicher für längerfristig vorzuhaltende digitale Objekte wissenschaftlichen Inhalts, in der Vergangenheit auch als Dokumentenserver oder Publikationsserver bezeichnet.

Von reinen Publikationsdatenbanken lassen sich IR über den Inhalt abgrenzen. Sie können neben den Metadaten auch das vollständige Dokument enthalten, neben klassischen

Publikationen können auch andere Objekte verzeichnet sein (etwa wissenschaftliche Primärdaten).

Von der Speicherung in CMS-Systemen unterscheidet sich ein IR über den Anspruch der dauerhaften Archivierung und Recherchierbarkeit wissenschaftlicher „Endprodukte“.

### **Grüner Weg**

Vor allem vor dem Hintergrund des „grünen Wegs“ im Rahmen der Diskussion für einen offenen Zugang zu wissenschaftlichem Wissen (Open Access) sind solche Systeme unverzichtbar.

Da es in Deutschland bereits ein gut ausgebautes Netz von Dissertationenservern gibt, können viele Bibliotheken bereits auf einschlägigen Erfahrungen aufbauen.

### **Exkurs: Goldener und grüner Weg**

Der "Grüne Weg" (Elektronische Zweitveröffentlichung)

Der sogenannte "Grüne Weg" beschreibt die Zweitveröffentlichung eines Artikels **neben** der Erstveröffentlichung in einer traditionellen Zeitschrift. Der Wissenschaftler publiziert in der Zeitschrift seiner Wahl und der Artikel wird in einem IR archiviert und damit auf einem zweiten Weg dem öffentlichen Zugang angeboten.

Normalerweise haben die meisten Autoren weitere Verwertungsrechte für ihren Artikel an die Verlage abgetreten. Aufgrund der langjährigen Diskussion haben die meisten Verlage Richtlinien veröffentlicht, die pauschal regeln, in welcher Form Autoren ihre Werke zweitveröffentlichen dürfen (obwohl sie im Vertrag die Rechte abgetreten haben). Gängige Einschränkungen hierbei ist, das nur die Nutzung des sogenannten „Final Draft“ (die Korrekturfahne nach der Peer Review) benutzt werden darf oder eine mehrmonatige Karenzzeit eingehalten werden muss. Das britische Sherpa-Projekt gibt in der Sherpa/Romeo-Liste hier hilfreiche Hinweise.

Diese Entwicklung schafft die Möglichkeit zum „Postprint“, d.h. die Zweitveröffentlichung der klassisch qualitätsgesicherten Fassung.

Der "Goldene Weg" (Open Access-Erstveröffentlichung)

In den letzten Jahren sind zahlreiche Open Access-Zeitschriften entstanden. Dies sind Zeitschriften, die auf der Basis verschiedenster Geschäftsmodelle die Erstveröffentlichung und gleichzeitigen offenen Zugang für wissenschaftliche Artikel ermöglichen. Dieses Modell ist im Handbuch bereits von Birgit Schmid als Thema des Quartals besprochen worden.

Für beide Modelle gilt: offener Zugang erhöht die Sichtbarkeit, erhöht den Impact. Auf Fragen zur Weiterentwicklung von Autorenverträgen (Stichworte wie Rechteevorbehalt, Creative (Science) Commons) kann an dieser Stelle nicht näher eingegangen werden.

## Technik

Technisch stützen sich IRs auf das Protokoll OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Basis des dort verwandten Schemas ist gängigerweise Dublin Core, es gibt aber auch Beispiele, bei denen andere Datenformate, wie MARC eingesetzt werden.

Unqualified Dublin Core:

DC:title  
DC:creator  
DC:date  
DC:subject  
DC:publisher  
DC:format  
DC:description  
DC:contributor  
DC:identifier  
DC:type  
DC:rights  
DC:language  
DC:source  
DC:relation  
DC:coverage

## OAI-PMH

Das Protokoll OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) macht das IR zum „data provider“, jeder Harvester („service provider“) kann die enthaltenen Daten maschinell abfragen („Transport“ in XML), mit Daten anderer „Data provider“ mischen, aufbereiten und zur Suche anbieten.

Ein Datensatz sieht (für den Harvester) beispielsweise so aus:

```
<record>
-<header>
<identifier>oai:HUBerlin.de:1354</identifier>
<timestamp>2005-01-26</timestamp>
<setSpec>dnb:01</setSpec>
<setSpec>pub-type:paper</setSpec>
</header>
-<metadata>
-<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Active networks and high speed content delivery</dc:title>
<dc:creator>Matyska, Ludek</dc:creator>
<dc:creator>Hladka, Eva</dc:creator>
<dc:subject>Wissenschaft und Kultur allgemein</dc:subject>
<dc:subject>active networks</dc:subject>
<dc:subject>active network model</dc:subject>
<dc:subject>programmable routers</dc:subject>
-<dc:description>
```

Future computer networks must be more flexible and faster than today. Active network paradigm is the way how to add flexibility to networks. During the last five years, a number of models for active networks realization were presented. In active networks, the concept of network as a distributed programmed facility is developed. This concept makes possible new applications and special services. Active network models are described and our active network model is described in more detail. We discuss applications which need special network features and the position active routers and switches have in network infrastructure in order to satisfy

these demands. An application which supports collaborative work, requiring data, audio and video streams to be transferred through the network is used to demonstrate how active network element can process dataflows with specific features in both uni- and multicast environments. An overly of active network elements placed in particular points of the non-intelligent high speed network is presented as a model which supports both the intelligence and high throughput without need of specialized costly hardware support.

```

</dc:description>
<dc:publisher>Humboldt University Berlin, Germany</dc:publisher>
<dc:date>2001-03-30</dc:date>
<dc:type>Text</dc:type>
<dc:type>paper</dc:type>
<dc:format>text/html</dc:format>
<dc:format>application/pdf</dc:format>
<dc:format>text/xml</dc:format>
<dc:language>eng</dc:language>
-<dc:identifier>
http://edoc.hu-berlin.de/eunis2001/d/Hladka/HTML/index.html
</dc:identifier>
-<dc:identifier>
http://edoc.hu-berlin.de/eunis2001/d/Hladka/PDF/Hladka.pdf
</dc:identifier>
-<dc:identifier>
http://edoc.hu-berlin.de/eunis2001/d/Hladka/XML/slides.xml
</dc:identifier>
</oai_dc:dc>
</metadata>
</record>

```

In der Aufbereitung für den Endnutzer kann dies sehr verschieden aussehen. Einige Beispiele:

Oaister:

<http://oaister.umdl.umich.edu>

The screenshot shows a Mozilla Firefox browser window titled "Oaister Search Results - Mozilla Firefox". The address bar contains the URL: <http://oaister.umdl.umich.edu/cgi/bib/bib-id?type=boolean&size=10&rgn1=title&rgn2=entire+record&rgn3=entire+>. The page displays the Oaister logo and navigation links (Home, Search, Browse Institutions, Help). A search summary indicates: "Your search was in the Title field for 'Active networks and high speed content delivery'. You found 1 record." Below this, a "Results by Institution" section lists "Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin" with 1 record. The main content area shows "Record 1 of 1" with a table of metadata:

Title	Active networks and high speed content delivery
Author/Creator	Matyska, Ludek; Hladka, Eva
Publisher	Humboldt University Berlin, Germany
Year	2001-03-30
Resource Type	Text
Resource Type	paper
Resource Format	text/html,application/pdf,text/xml
Language	English
Note	Future computer networks must be more flexible and faster than today. Active network paradigm is the way how to add flexibility to networks. During the last five years, a number of models for active networks realization were presented. In active networks, the concept of network as a distributed programmed facility is developed. This concept makes possible new applications and special services. Active network models are described and our active network model is described in more detail. We discuss applications which need special network features and the position active routers

The browser's taskbar at the bottom shows the Start button, a taskbar with "DE", "Microsoft Word", and "Oaister Search Resu...", and a system tray with the time "08:29".

Base (Bielefeld Academic Search Engine):

<http://digital.ub.uni-bielefeld.de/>

The screenshot shows the Base search engine interface. The browser window title is "BASE - Bielefeld Academic Search Engine - Mozilla Firefox". The address bar shows the URL [http://digital.ub.uni-bielefeld.de/index.php?BASE\\_DigitalCollections=6b7524d5a196ab27ca01876eceb9a3e9q=Active4](http://digital.ub.uni-bielefeld.de/index.php?BASE_DigitalCollections=6b7524d5a196ab27ca01876eceb9a3e9q=Active4). The search results for "Active networks and high speed content delivery" are displayed, showing 148 documents found in 0.1948 seconds. The first result is titled "Active networks and high speed content delivery" by Matyska, Luděk and Hladka, Eva. The second result is titled "Scalable Multiple Description Coding and Distributed Video Streaming over 3G Mobile Networks" by Zheng, Ruobin. The interface includes a search bar, a "Suchen!" button, and various filters for sorting and refining results.

Google Scholar:

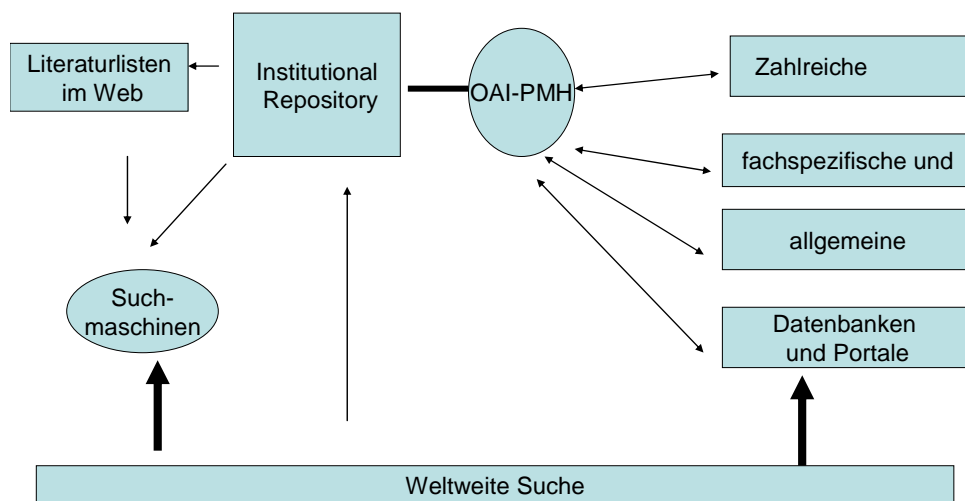
[<http://scholar.google.de>](http://scholar.google.de)

The screenshot shows the Google Scholar search results page. The browser window title is "Active networks and high speed content delivery" - Google Scholar - Mozilla Firefox. The address bar shows the URL <http://scholar.google.de/scholar?hl=de&lr=&q=%22Active+networks+and+high+speed+content+delivery%22&btnG=>. The search results for "Active networks and high speed content delivery" are displayed, showing 1 result found in 0.07 seconds. The result is titled "Active networks and high speed content delivery" by Hladka, L. and Matyska, L. The interface includes a search bar, a "Suche" button, and a "Web-Suche" option.

## Dissemination

Institutional Repositories werden also mit ihren Suchoberflächen zwar durchaus lokal in einer Institution genutzt. Ihr entscheidender Wert liegt aber darin, dass die Daten maschinell lesbar für Harvester zur Verfügung gestellt werden. Nicht nur die Inhalte sind „Open Access“, auch das Repository ist offen zugänglich und dient mit seiner Technik der Dissemination der enthaltenen wissenschaftlichen Ergebnisse.

Auf längere Sicht entsteht eine Landschaft sich gegenseitig bedienender Dienste.



## Risiko

Im Sinn einer Risikoanalyse sind vor Einführung eines IR verschiedene Aspekte zu betrachten. Bedenkenswert sind dabei vor allem zwei Punkte: der Aufbau eines solchen Systems verursacht dauerhaft Kosten, die kalkuliert werden müssen. Entscheidender ist aber der Hinweis, dass ein solches System von der technischen Seite schnell erstellt ist, ausschlaggebend für den Erfolg aber die Sammlung von Inhalten ist. Alle Erfahrungen gehen davon aus, dass dieser Erfolg nur durch Stützung durch entsprechende Mandate in den Institutionen (Policies) erreicht werden kann.

<i>Ereignis</i>	<i>Wahrscheinlichkeit</i>
Alle Universitäten und Forschungseinrichtungen haben IRs – kurzfristige Perspektive	Niedrig
Alle Universitäten und Forschungseinrichtungen haben IRs – langfristige Perspektive	Sehr hoch
Keine Universität und Forschungseinrichtung hat IR	Niedrig
IR ist leer	Hoch
Dort dokumentierte Forschung wird nicht wahrgenommen	Sehr niedrig

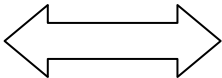
Es entsteht kein Nutzen	Sehr niedrig
Zu hohe Einrichtungskosten	Mittel
Zukünftig hohe Kosten	Sehr niedrig
Urheberrechtsprobleme	Sehr niedrig
Wissenschaftler wollen nicht zusammenarbeiten	Sehr niedrig

Nach A. Sale: Generic Risk Analysis - Open Access for your institution

### Workflows

Für den erfolgreichen Betrieb eines Repositories müssen Workflows erarbeitet werden, die die Strukturen der Institution berücksichtigen und der Bibliothek ein effektives Arbeiten möglich machen. Die technischen Systeme hinter den Repositories müssen diese dann so einfach wie möglich abbilden.

Entscheidend ist hier die Bestimmung der eigenen Position der Bibliothek im Spannungsverhältnis zwischen „Selbstarchivierung durch den Autor“ und „Rund um Sorglos-Paket“.

<i>Selbstarchivierung Autor</i>		<i>Sorglos-Paket</i>
Geringer Aufwand für Bibliothek		Hoher Aufwand
Risiko: leeres Repository		Gefülltes, funktionierendes Rep.

### Systemwahl

Für eine Systemwahl gelten alle klassischen Kriterien wie in Kapitel 9.3 dargelegt. Da viele Bibliotheken aber schon traditionell Verzeichnisse der Publikationen einer Institution führen (Publikationsdatenbank), bietet sich an bei der Einführung von IRs entsprechende Workflows zu nutzen und weiter auszubauen. Da IRs den direkten Zugang zum Volltext ermöglichen und nicht nur Metadaten verzeichnen, wird im ersten Schritt zu prüfen sein, inwieweit die vorhandenen Systeme erweiterbar oder mit IRs verknüpft werden können.

Weiterhin wird zu prüfen sein, inwieweit ein solches System in bereits vorhandene oder geplante Dienstplattformen der Institution integrierbar und mit anderen Diensten verknüpfbar ist. Webservices auf der Basis von SOAP spielen hier eine unverzichtbare Rolle. Verbünde bieten inzwischen das Hosting für IRs an, auch hier ist das für und wider einer lokalen Installation zu prüfen.

Im Sinn von „Subject Repositories“, d.h. also Archivierung und Verbreitung nicht vor einem institutionellen Hintergrund, sondern auf der Basis eines Fachkontextes werden von einer Reihe von Virtuellen Fachbibliotheken entsprechende Möglichkeiten angeboten. Dies scheint besonders für kleine Institutionen eine sehr gangbare Perspektive zu sein.

Elementare Punkte bei der Auswahl von IRs:

- OAI-PMH
- Webinterface
- Umgang mit Datenspeicherung (Perspektive: Langzeitarchivierung)
- Abbildung von Workflows
- Interne Rechteverwaltung
- Persistente Identifier
- Webservices

## **Systeme**

Neben Eigenentwicklungen werden zur Zeit u.a. folgende Systeme häufig eingesetzt:

- CDSware <http://cdsware.cern.ch/> (Das System des CERN. Das CERN ist Vorreiter in Sachen „grüner Weg“)
- Eprints <http://www.eprints.org/> (das am meisten verbreitete System in Grossbritannien).
- Dspace <http://www.dspace.org/>
- Fedora <http://www.fedora.info/> (mehr eine Toolbox, als ein System “Out of the box”).
- Mycore <http://www.mycore.de/projektbeschreibung.html>
- Opus <http://opus.uni-stuttgart.de/opus/sw> (hervorgegangen aus einer Entwicklung für Dissertationenserver, längst ein vollwertig einsetzbares IR).

Systeme wie Digitool oder ContentDM haben Komponenten, die sie vor allem für die Verwaltung digitaler Objekte in einem breiten Sinn (Bilder udgl.) interessant machen.

Interessant wird sein, die Entwicklung von eSciDoc (Max-Planck-Gesellschaft zusammen mit FIZ Karlsruhe) zu beobachten. Hier soll, finanziert vom BMBF, ein System entwickelt werden, dass viele Anforderungen befriedigen kann. Als Plattform hat man sich für Fedora entschieden, mit einer nachnutzbaren Applikation ist aber wohl nicht vor Ende 2007 zu rechnen.

Eine umfassende Darstellung zum gesamten Komplex Institutional Repositories findet sich im 2006 erschienenen Buch von Jones u.a.: The Institutional Repository.

## **Ausblick**

Über die vielfältige Verknüpfung der Repositories durch Harvester entsteht eine neue, vielfältig verknüpfbare Landschaft von Nachweisen, die direkten Zugang zu Volltexten als integralen Bestandteil einer sich entwickelnden E-Science ermöglichen.

Dieses Netz wird ergänzt werden durch weitere ‚Objekte‘ in den Repositories, die Publikationen ergänzen: von wissenschaftliche Daten über Bilder zu bewegten Visualisierungen.

Hinweise:

Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen (2003).  
<[http://www.mpg.de/pdf/openaccess/BerlinDeclaration\\_dt.pdf](http://www.mpg.de/pdf/openaccess/BerlinDeclaration_dt.pdf)>

Elektronisches Publizieren an Hochschulen : Inhaltliche Gestaltung der OAI-Schnittstelle ; Empfehlungen (2005). DINI Arbeitsgruppe Elektronisches Publizieren.  
<<http://edoc.hu-berlin.de/series/dini-schriften/2005-2-de2/PDF/2-de2.pdf> >



Empfehlungen zur digitalen Informationsversorgung durch Hochschulbibliotheken /  
Wissenschaftsrat (2001)

<<http://www.wissenschaftsrat.de/texte/4935-01.pdf>>

Foster, N. F.; Gibbons, S. (2005): Understanding faculty to improve content recruitment for  
institutional repositories. D-Lib Magazine, 11,1.

<<http://www.dlib.org/dlib/january05/foster/01foster.html>>

Barton, M. R., Waters, M. M. (2005): Creating an Institutional Repository:  
LEADIRS Workbook. – MIT Libraries.

<<http://www.dspace.org/implement/leadirs.pdf>>

Guide to institutional repository software.

<<http://www.soros.org/openaccess/software/>>

Guidelines for Repository Implementers.

<<http://www.openarchives.org/OAI/2.0/guidelines-repository.htm>>

JISC: Digital Repositories (2005).

<[http://www.jisc.ac.uk/uploaded\\_documents/HE\\_repositories\\_briefing\\_paper\\_2005.pdf](http://www.jisc.ac.uk/uploaded_documents/HE_repositories_briefing_paper_2005.pdf)>

Jones, R., Andrew, T., MacColl, J. : The Institutional Repository. Oxford : Chandos, 2006. -  
247 S.

ISBN 1-84334-138-7

<<http://hdl.handle.net/1842/858>>

OAI for beginners.

<<http://www.oaforum.org/tutorial/>>

Sale, A (2006): General Risk Analysis : Open Access for your Institution.

<<http://eprints.comp.utas.edu.au:81/archive/00000266/>>

Sherpa: Publisher copyright policies & self-archiving: the SHERPA/ROMEO list.

<<http://www.sherpa.ac.uk/romeo.php>>

SPARC Repository Resources.

<<http://www.arl.org/sparc/repos/> >

SPARC Institutional Repository Checklist and Resource Guide.

<[http://www.arl.org/sparc/IR/IR\\_Guide\\_v1.pdf](http://www.arl.org/sparc/IR/IR_Guide_v1.pdf)>